

( Translation )

PATENT OFFICE  
JAPANESE GOVERNMENT

This is to certify that the annexed is a true copy of the  
following application as filed with this Office

Date of Application	March 28, 2000
Application Number	Japanese Patent Application No 088695 2000
Applicant(s)	Hitachi Software Engineering Co Ltd

August 18, 2000

Commissioner,  
Patent Office

Kozo Oikawa (seal)

Certificate No 2000-3066087



日 本 国 特 許 庁  
PATENT OFFICE  
JAPANESE GOVERNMENT

JCS11 U.S. PTO

09/677042



別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日

Date of Application:

2000年 3月28日

出 願 番 号

Application Number:

特願2000-088695

出 願 人

Applicant (s):

日立ソフトウェアエンジニアリング株式会社

2000年 8月18日

特許庁長官  
Commissioner,  
Patent Office

及川耕造

出証番号 出証特2000-3066087

【書類名】 特許願  
【整理番号】 SK11A117  
【提出日】 平成12年 3月28日  
【あて先】 特許庁長官殿  
【国際特許分類】 G06F 19/00

【発明者】

【住所又は居所】 神奈川県横浜市中区尾上町6丁目81番地 日立ソフト  
ウェアエンジニアリング株式会社内

【氏名】 中重 亮

【発明者】

【住所又は居所】 神奈川県横浜市中区尾上町6丁目81番地 日立ソフト  
ウェアエンジニアリング株式会社内

【氏名】 野崎 康行

【発明者】

【住所又は居所】 神奈川県横浜市中区尾上町6丁目81番地 日立ソフト  
ウェアエンジニアリング株式会社内

【氏名】 渡辺 恒彦

【発明者】

【住所又は居所】 神奈川県横浜市中区尾上町6丁目81番地 日立ソフト  
ウェアエンジニアリング株式会社内

【氏名】 田村 卓郎

【特許出願人】

【識別番号】 000233055

【氏名又は名称】 日立ソフトウェアエンジニアリング株式会社

【代理人】

【識別番号】 100083552

【弁理士】

【氏名又は名称】 秋田 収喜

【電話番号】 03-3893-6221

【手数料の表示】

【予納台帳番号】 014579

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 遺伝子発現パターン表示方法および装置並びに記録媒体

【特許請求の範囲】

【請求項 1】 複数の遺伝子の発現パターンを視覚的に表示する遺伝子発現パターン表示方法であって、

前記遺伝子の発現パターンのデータをクラスタ分析した結果に対し、識別誤差範囲を考慮した段階別クラスタ概数を表示するステップを備えることを特徴とする遺伝子発現パターン表示方法。

【請求項 2】 前記クラスタ分析の結果において、あるクラスタに属する遺伝子発現パターンデータの個数が所定数以上の場合に限り、遺伝子グループとしてグループ化して表示することを特徴とする請求項 1 記載の遺伝子発現パターン表示方法。

【請求項 3】 複数の遺伝子の発現パターンを視覚的に表示する装置であって、

前記遺伝子の発現パターンのデータを格納した記憶手段と、記憶された遺伝子パターンのデータを読み出し、クラスタ分析する分析処理手段と、分析結果に対し、識別誤差範囲を考慮した段階別クラスタ概数を表示装置に表示する表示処理手段とを備えることを特徴とする遺伝子発現パターン表示装置。

【請求項 4】 表示装置画面に複数の遺伝子の発現パターンを視覚的に表示する処理プログラムを記録した媒体であって、

前記遺伝子の発現パターンのデータをクラスタ分析した結果に対し、識別誤差範囲を考慮した段階別クラスタ概数を表示するステップを含む処理プログラムが記録されていることを特徴とするコンピュータが読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、特定の遺伝子とハイブリダイズさせることによって得られた時系列の遺伝子発現パターンデータを視覚的に分かり易く、そして遺伝子の機能・役割が推測し易い形式で表示するための遺伝子発現パターン表示方法および装置に

関するものである。

【0002】

【従来の技術】

従来、ゲノム配列が決定された種の増加に伴い、進化に対応すると見られる遺伝子を見つけ出し、どの生物にも共通に持っていると考えられる遺伝子の集合を探したり、それから逆に種に個別な特徴を推測するなど、種間の遺伝子の違いから何かを見出そうとする、いわゆるゲノム比較法が盛んに行われてきた。

【0003】

しかし近年、DNAチップやDNAマイクロアレイなどのインフラストラクチャの発達によって、分子生物学の興味は、種間の情報から種内の情報へ、すなわち同時発生解析へと移りつつあり、これまでの種間の比較と合わせて、情報の抽出から関連付けの場が大きく広がりを持ち始めている。

【0004】

例えば、既知の遺伝子と同一の発現パターンを示す未知の遺伝子が見つければ、それが既知の遺伝子と同様の機能があると類推できる。これら遺伝子や蛋白質そのものの機能的な意味付けは、機能ユニットや機能グループといった形で研究されている。また、それらの間の相互作用も、既知の酵素反応データや物質代謝データとの対応付けによって、あるいはより直接的に、ある遺伝子を破壊あるいは過剰反応させ、その遺伝子の発現をなくすか、あるいは多量に発現させ、その遺伝子の直接的および間接的影響を、全遺伝子の発現パターンを調べることによって解析している。

【0005】

この分野において成功した事例として、スタンフォード大学のP. Brownらのグループによるイースト菌の発現解析が挙げられる (Michel B. Eisen et al. : Cluster analysis and display of genome-wide expression patterns : Proc. Natl. Acad. Sci. (1998) Dec 8 ; 95(25) : 14863-8)。彼らは、DNAマイクロアレイを用いて、細胞から抽出した遺伝子を時系列にハイブリダイズさせ、遺伝子の発現の度合い (ハイブリダイズした蛍光シグナルの輝度) を数値化した。数値に色を対応させることで、遺伝子の個々の発現過程を分かり易く表示させている

。このとき、細胞の一連のサイクルにおいて発現パターンの過程が近い遺伝子同士（任意の時点での発現の度合いが近いもの同士）をクラスタリングしている。

#### 【0006】

図12は、この方法にそって遺伝子の発現状態1200を表示した例であり、横方向に実験ケース、縦方向に遺伝子を並べている。また左側の樹状図は、クラスタリングの過程で、最も近い2つのクラスタ毎に併合されてきた状況を表しており、各枝の長さは併合時の2つのクラスタ間距離に対応している。なお、図12における1つの枠1201が1つの遺伝子のある時刻における発現状態を示すものであり、図12では白黒の濃度を変えて発現状態を模式的に示している。

このような表示方法をとることにより、共通のクラスタに属する遺伝子は、共通の機能的性質をもつ可能性がある」と類推することができる。

#### 【0007】

##### 【発明が解決しようとする課題】

ところが、実際の遺伝子発現パターンの分析では、図12と同様な大量のデータをクラスタリングすることになる。遺伝子の種類に関しては、数千から一万、最大では十万を超える量であり、実験ケースも10程度のオーダーから数十、数百など、任意の個数のデータを用意することが出来る。このため、図12の樹状図の部分も非常に複雑な、細かな枝を多量に含んだものになる。

この状況を表わしたものが図13である。大量の遺伝子発現パターンデータを対象にしたクラスタリングの結果全体が図13の左側の部分である。また、右側の点線1301で囲んだ部分は、結果全体のうちユーザが実際に注目して分析結果の詳細を見るため、ウィンドウ枠などで範囲を限定した状況を示している。

このようにして得られた樹状図1302は、クラスタの最も近いものを2つずつ併合してきた過程を正確に表わしているが、この表示を見て遺伝子のグループ分けを判断・推測するユーザが見て、どのくらいのクラスタ数で大まかな分類ができていのかを判断するのは難しい、という難点がある。

ユーザは、例えば十くらい、百くらいの大まかな分類の仕方を提示してもらえらるほうがありがたい。すなわち、大きな差異がある分類段階を自動的に計算し、例えば7クラスの場合、28クラスの場合、105クラスの場合、372クラス

の場合など、メニューとしての提示があれば、ユーザは細かな差異によるクラスタリングの影響に煩わされることなく、大まかな粒度に応じた適切な分類結果を選択して、遺伝子のグループ分けを考察することができる。

#### 【0008】

本発明は、このような従来技術の問題点を鑑み、クラスタリングの結果から、より大まかな分類結果を自動的に抽出し、ユーザが適切な分類結果を選択して分かり易く表示し、遺伝子のグループ分けを考察することができるようにする遺伝子発現パターン表示方法および装置を提供することを目的とする。換言すれば、上述したクラスタリングの過程で、分類間の隔たりが大きくなるようなクラスタリング結果として、複数の段階を用意し、効果的に表示することができる遺伝子発現パターン表示方法および装置を提供することを目的とする。

#### 【0009】

##### 【課題を解決するための手段】

本発明では、前記目的を達成するために、遺伝子の発現パターンのデータをクラスタ分析した結果に対し、識別誤差範囲を考慮した段階別クラスタ概数を表示するステップを備えることを特徴とする。

例えば、クラスタリング処理過程で、ユーザの指定した識別誤差範囲を超える分類結果を複数保持しておき、結果の表示において、ユーザへ大まかな分類結果を複数提示する。

結果の表示においては、複数の分類結果の中からある分類を選択するための区間スケールバーと樹状図の切断線を用意する。ユーザは、区間スケールバー上の樹状図切断線を動かすことによって、特定の分類を選択できる。樹状図切断線を動かす際には、その位置での分類結果で幾つかのクラスタに分かれているかを明示するために、その遺伝子グループ数を表示する。また、遺伝子発現パターンデータの表示部分には、その樹状図切断線位置での遺伝子グループ間の境界線を引き、一定規模以上の遺伝子グループを明示的に強調表示する。

#### 【0010】

##### 【発明の実施の形態】

以下、図面を参照して本発明の実施の形態を説明する。

図1は本発明の遺伝子発現パターン表示方法を適用した遺伝子発現パターン解析装置の一実施形態を示すシステム構成図である。この実施形態の解析装置は、一連の細胞のプロセスにおいて遺伝子の発現の度合いを数値化した遺伝子発現パターンデータ100を格納している記憶装置（またはデータベース）101、その発現パターンデータを視覚化して表示するための表示装置102、本システムへの値の入力や選択の操作を行うためのキーボード103、マウス104、遺伝子の発現過程に応じて発現パターンデータ100のクラスタリングを行なう表示処理部105から構成される。表示処理部105は、コンピュータと表示のための処理を行うプログラムによって具体化されるものである。なお、記憶装置101は、ネットワークを介して表示処理部105と結合することも可能である。

#### 【0011】

図2は、本発明によるクラスタ分析結果の表示例である。

図2は、大量の遺伝子発現パターンデータを対象にしたクラスタリングの結果全体を示した図13左側部分の一部を基に表示したものであり、図13の右側の点線1301で囲まれた部分に対応している。もちろん、本発明は、図2自体がクラスタリング結果の全体となり、図13の左側と図2が一致する場合も含むものである。

図2では、区間スケールバー201上を動かことが可能な樹状図切断線202が、ユーザによるマウス104などのポインティングデバイスを利用した指示により、区間スケールバー201上の第2区間(②)に置かれた状況を示している。第2区間上では、この区間での分類によって遺伝子グループが3個出来ているという情報(クラスタ概数)203を示している。

さらに、遺伝子発現パターンデータ200の部分には、この分類に対応した境界線204を引き、遺伝子グループA、B、Cとして分類結果に対する強調表示205を行っている。

#### 【0012】

図3は、記憶装置101に格納された遺伝子発現パターンデータ100の具体的な構造を示したものである。ここで例示する遺伝子発現パターンデータ100には、各遺伝子の遺伝子ID(geneID)301に対応してm個のベクトルデータ

3 0 2 がある。配列のインデックスは  $n$  個の実験ケースに対応し、配列要素の中身には、遺伝子の発現の度合い（ハイブリダイズした蛍光シグナルの輝度）を数値化したデータを格納している。

#### 【 0 0 1 3 】

図 4 はクラスタリング処理において利用するクラスタ構造体の例を示すものである。クラスタ構造体には 2 種類あり、typeメンバの値がleafのもの（左側）4 0 1 とnodeのもの（右側）4 0 2 に分かれる。

leaf型クラスタ構造体 4 0 1 は、各遺伝子ごとの発現パターンデータ、すなわち、図 3 の各配列データに対応するもので、その遺伝子 ID の値をgeneIDメンバの値（例えば 1 7）として設定する。また、クラスタとしてのlevelはゼロに設定する。

node型クラスタ構造体 4 0 2 はクラスタリングにおける併合処理において逐次生成するもので、併合前の 2 つのクラスタをleftメンバの値とrightメンバの値から迎れるようにし、また、それらの間の距離をdistanceメンバの値として保持する。クラスタとしてのlevelは、識別誤差範囲の値に応じて（例えば 5）設定する。

#### 【 0 0 1 4 】

図 5 は、クラスタ分析の過程で生成するデータ構造を示した図である。クラスタ構造体は、最初leaf型の構造体 4 0 1 だけを用意するが、クラスタ分析の過程で 2 つずつ併合し、その度にnode型クラスタ構造体 4 0 2 を生成してトリートメントを組み立てる。node型クラスタ構造体 4 0 2 はそれを生成した順に、逐次、配列node\_clusters[] から迎れるようにポインタを張ってゆく。変数nclus 5 0 1 は、これまで生成したnode型クラスタ構造体 4 0 2 の総数を保持する変数である。

#### 【 0 0 1 5 】

図 6 は、区間スケールバー 2 0 1 や、ある分類結果における遺伝子グループ数の情報 2 0 3、遺伝子グループ間の境界線 2 0 4 を引くための表示情報を保持するためのデータ構造例を示した図である。区間スケールバー 2 0 1 上の区間に対応するinterval構造体 6 0 1 は、識別誤差範囲を考慮した分類決定処理の過程で逐次生成されて、配列intervals[] 6 0 2 から迎れるようにポインタを張ってゆ

く。変数 level は、これまで生成した interval 構造体の総数を保持する変数である。

各 interval 構造体 6 0 1 は、メンバ名として num\_clust、min\_dist、max\_dist、borders があり、num\_clust メンバには、その分類における遺伝子グループ数情報が、min\_dist メンバと max\_dist メンバには、クラスタ間距離に関する区間の上限値と下限値が、borders メンバには、遺伝子発現パターンデータ 2 0 0 中の境界線を引く位置（行番号）を設定する。

また、配列 disp\_leaf\_clusters [] 6 0 3 には、遺伝子発現パターンデータ 2 0 0 の各行に、図 3 中のどの遺伝子 ID に対応するベクトルデータを表示するかを決定するため、第 i 行に対応する第 i 番目の配列要素として表示する遺伝子 ID の値を設定する。

【 0 0 1 6 】

図 7 は、本発明の遺伝子発現パターン表示方法における概略処理手順を示すフローチャートである。

まず、記憶装置 1 0 1 に格納された遺伝子発現パターンデータ 1 0 0 を表示処理部 1 0 5 へ読み込む（ステップ 7 0 0）。この場合、個々の発現パターンデータ 1 0 0 の具体的な構造は、図 3 に示したものである。

次に、クラスタ分析に必要な各種パラメータを設定する（ステップ 7 0 1）。この設定段階で、キーボード 1 0 3 を用いたユーザから識別誤差範囲の値の入力を受付け、変数 E に保持する。この識別誤差範囲の値は、図 3 に示した遺伝子ごとの発現パターンデータを 2 つずつ、距離や非類似度、類似度などの尺度に基いて比較する際、この値以上の差異があれば別のデータとして識別すべきである、という閾値を意味している。

各種パラメータ設定の後、クラスタ分析を行う（ステップ 7 0 2）。このクラスタ分析の処理の間に、本発明の表示に必要な情報を収集し、表示用データの計算を行う。これについては、後で詳しく説明する。

最後に分析結果の表示を行う（7 0 3）。ここで、先に収集し、計算しておいた表示用のデータを用い、本発明に特有な表示（図 2 における区間スケールバー 2 0 1、樹状図切断線 2 0 2、遺伝子発現パターンデータ中の境界線 2 0 4、遺

伝子グループの強調表示 205) を行う。

ここで、遺伝子グループの強調表示 205 を行う際には、所定数以上の遺伝子発現データが 1 つのクラスタとしてまとめられている場合に限って表示し、少数の遺伝子発現データで構成されたクラスタに対しては、遺伝子グループとしての表示を行わない、などの方法も可能である。

#### 【0017】

図 8 は、図 7 におけるクラスタ分析 (ステップ 702) の処理の詳細を示すフローチャートであり、第一段階として実行するクラスタ木の生成処理に関するフローチャートである。

図 8 において、まず、図 3 に示した各遺伝子 ID 301 に対応する  $m$  個のベクトルデータ 302 を  $m$  個の leaf 型クラスタ構造体 401 とし、併合対象クラスタとして登録する (ステップ 800)。次に、併合対象クラスタ数  $cnum$  の値を  $m$ 、これまで生成した node 型クラスタ構造体 402 の数  $nclus$  を「0」として初期化する (ステップ 801)。さらに、併合対象クラスタの数  $cnum$  が「1」に等しいかどうか判定し (ステップ 802)、等しくない場合、「1」になるまで以下の一連の処理を繰り返す。

#### 【0018】

最初に、登録された併合対象のクラスタ構造体から相対距離最小の 2 つのクラスタを選択する (ステップ 803)。次に、node 型クラスタ構造体 C を新規に生成し (ステップ 804)、node 型クラスタ数をインクリメントする (ステップ 805)。そして、配列 `node_clusters[]` の第  $nclus$  番成分に新しい node 型クラスタ構造体を登録する (ステップ 806)。さらに、新しい node 型クラスタ構造体の left メンバ、right メンバ、distance メンバに、先にステップ 803 で選択した 2 つのクラスタ、およびその間の距離を登録する (ステップ 807)。

ここで、2 つクラスタのどちらを left メンバとし、残りを right メンバとするかについて、予め判定基準を設ける方法を採用することも可能である。

最後に、この 2 つのクラスタ構造体を併合対象クラスタ構造体から除外、新しい node 型クラスタ構造体を登録し (ステップ 808)、併合対象クラスタ数  $cnum$  の値をデクリメントする (ステップ 809)。

ステップ 8 0 2 の判定において `cnum` の値が「1」に等しくなった場合は、図 9 のフローチャートに示す処理（クラスタレベルの設定）に継続する。

#### 【0 0 1 9】

図 9 は、第 2 段階として実行するクラスタレベルの設定処理に関するフローチャートである。

第 1 段階の処理で生成された `node` 型クラスタ構造体はすべて配列 `node_clusters[]`（図 5）に登録されており、その `distance` メンバの値は配列のインデックスに従って一般に昇順データをなっているが、クラスタリングのアルゴリズムの選択によっては、必ずしも昇順とはならない可能性がある。このため、まず、配列 `node_clusters[]` に登録された各 `node` 型クラスタ構造体の `distance` メンバを検証する（ステップ 9 0 0）。

この検証の結果、昇順になっていない場合、`node_clusters[]` を昇順になるようにソート処理を施す、あるいは昇順になっていない場所だけ検出して別の処理を行う。

#### 【0 0 2 0】

次に、各種変数の初期値を設定する（ステップ 9 0 1）。具体的には、配列 `node_clusters[]` 用カウンタ `i`（初期値 1）、クラスタレベル `level`（初期値 0）、処理中の `node` 型クラスタの `distance` メンバ値 `curr_dist`（初期値 0）、前回処理した `node` 型クラスタの `distance` メンバ値 `prev_dist`（初期値 0）を設定する。

そして、カウンタ `i` の値と変数 `nclus` の値を比較することにより、配列 `node_clusters[]` の各要素に対して、以下の一連の処理を実行する（ステップ 9 0 2）。

#### 【0 0 2 1】

まず、`i` 番目の `node` 型クラスタの `distance` メンバ値を変数 `curr_dist` に保持し（ステップ 9 0 3）、変数 `curr_dist` の値と `prev_dist` の値の差がユーザにより指定された識別誤差範囲 `E` の値より小さいかどうかを判定する（ステップ 9 0 4）。

その差が `E` の値より小さくない場合は、まず変数 `level` の値をインクリメントする（ステップ 9 0 5）。そして、新しく `interval` 構造体を生成し、その `min_dist` メンバに `prev_dist` の値を、`max_dist` メンバに `curr_dist` の値を、`num_clust` メ

ンバに  $(nclus - i + 1)$  の値を設定し、このinterval構造体自体を配列 `intervals[]` の第level番目の要素として登録する（ステップ906）。ここで、`borders` メンバにはデフォルト値として空集合  $\{\}$  を設定しておく。

#### 【0022】

ステップ904の判定において、`curr_dist`の値と`prev_dist`の値の差がEの値より小さい場合には、ステップ905とステップ906の処理を省略し、以下の処理に継続する。

また、`node_clusters[i]`に登録されたnode型クラスタ構造体のlevelメンバ値として、変数levelの値を登録する（ステップ907）。

最後に、`curr_dist`の値を`prev_dist`に移し、カウンタiの値をインクリメントして（ステップ908）、ステップ902に戻る。

配列`node_clusters[]`のすべての要素に対して、ステップ903からステップ908までの処理を実行し終えたら（ステップ902）、図10の処理に継続する。

#### 【0023】

図10は、図7におけるクラスタ分析（ステップ702）の処理の詳細を示すフローチャートであり、第3段階として実行する表示用データの作成処理に関するフローチャートである。

まず、配列`disp_leaf_clusters[]`のインデックスを保持する変数jの値を「1」とし、次の表示用データ作成処理（処理A）に渡す引数`cluster`の値を、配列`node_clusters[]`の第nclus番目の要素が指すnode型クラスタ構造体として設定する（ステップ1000）。

そして、`cluster`を引数として、処理A：表示用データ作成処理ルーチンを呼び出す（ステップ1001）。このステップ1001では、処理Aを再帰的に呼び出し、図8のフローで作成したクラスタ木の構造に従ってトリーウォークを実行する過程で、表示用のデータを収集し、計算している。この再帰的な処理が終了した段階で、クラスタ分析の処理を終了する。

#### 【0024】

図11は、図10における処理A：表示用データ作成処理（ステップ1001

) の詳細を示すフローチャートである。

まず、引数culsterで渡されたクラスタ構造体のtypeメンバの値を検査する（ステップ1100）。

その結果、leaf型クラスタ構造体であった場合、配列disp\_leaf\_clusters[]の第j番目の要素として、そのクラスタのgeneIDメンバの値を設定し（ステップ1101）、jの値をインクリメントする（ステップ1102）。

#### 【0025】

また、引数clusterで渡されたクラスタがnode型クラスタ構造体であった場合、まずleftメンバのクラスタを引数とした処理Aの再起呼び出し（ステップ1103）を行い、次に自分のクラスタに関する処理（ステップ1104からステップ1106まで）を行い、最後にrightメンバのクラスタを引数とした処理Aの再起呼び出し（ステップ1107）を行う。

#### 【0026】

自分のクラスタに関する処理としては、まず、自分のlevelメンバ値、leftメンバから迎れるクラスタのlevelメンバ値、rightメンバから迎れるクラスタのlevelメンバ値の3データをそれぞれ変数my\_level、l\_level、r\_levelに設定する（ステップ1104）。次にl\_levelとr\_levelの小さいほうの値と、my\_levelの値を比較する（ステップ1105）。

その結果、my\_levelの方が大きかった場合は、 $\min\{l\_level, r\_level\}$  から(my\_level -1)までインデックスkの値を動かし、intervals[k]の指すinterval構造体のbordersメンバに対して、その値の集合に変数jの値を追加する処理を行う（ステップ1106）。

ステップ1105での比較結果、my\_levelの方が大きくなかった場合は、ステップ1106の処理を省略し、ステップ1107に移る。

以上、leaf型クラスタ構造体またはnode型クラスタ構造体に対する一連の処理が終了した場合、処理Aは終了する。

#### 【0027】

以上の処理によって、図2に示したようなクラスタ分析結果の表示が可能となる。

まず、遺伝子発現パターンデータ 200 の部分は配列 `disp_leaf_clusters[]` の情報を利用することによって、上から 1 行ずつ表示できる。また樹状図の部分は、`node_clusters[nclus]` の指しているクラスタを根として持つトリートリー構造のデータから表示が可能である。

本発明の特徴となる区間スケールバー 201 と樹状図切断線 202 は、配列 `intervals[]` に登録された各 `interval` の `min_dist` メンバ値と `max_dist` メンバ値を参照することで表示可能であり、分類結果の遺伝子グループ数の表示 203 は、`num_clust` メンバ値を参照することで表示可能である。

さらに、遺伝子発現パターンデータの表示中にある遺伝子グループ間の境界線 204 とグループに対する範囲の強調表示 205 は、`borders` メンバ値を参照することで表示可能である。

【0028】

なお、図 7～図 11 に示した処理は、CD-ROM 等の記録媒体に遺伝子発現パターン解析表示プログラムとして記録してパーソナルコンピュータやワークステーション等のユーザに提供することができる。

【0029】

【発明の効果】

以上説明したように、本発明によれば、クラスタリングの結果から、より大まかな分類結果を自動的に抽出し、その中からユーザが所望する段階の分類結果を選択して分かり易く表示することができる。すなわち、分類間の隔たりが大きくなるようなクラスタリング結果として複数の段階を用意し、効果的に表示することができる。したがって、ユーザは遺伝子のグループ分けを判断・推測する際にこの表示を見て、どのくらいのクラスタ数で大まかな分類ができているのかを容易に判断することができるようになる。

【図面の簡単な説明】

【図 1】

本発明を適用した遺伝子発現パターン解析装置の一実施形態を示すシステム構成図である。

【図 2】

遺伝子発現パターンに対する本発明のクラスタ分析結果の表示例を示す図である。

【図 3】

遺伝子発現パターンデータの構造例を示す図である。

【図 4】

クラスタ構造体の例を示す図である。

【図 5】

クラスタ木構造の生成例を示す図である。

【図 6】

表示用データの例を示す図である。

【図 7】

本発明の遺伝子発現パターンの表示処理の概略処理手順を示すフローチャートである。

【図 8】

クラスタ分析の中でクラスタ木の生成処理を示すフローチャートである。

【図 9】

クラスタ分析の中でクラスタレベルの設定処理を示すフローチャートである。

【図 10】

クラスタ分析の中で表示用データ作成処理を示すフローチャートである。

【図 11】

表示用データ作成処理の中で処理 A の詳細を示すフローチャートである。

【図 12】

遺伝子発現パターンに対する標準的クラスタ分析結果の表示例を示す図である。

【図 13】

クラスタ分析結果の全体と表示対象部分木の対応例を示す図である。

【符号の説明】

100…遺伝子発現パターンデータ、101…記憶装置、102…表示装置、  
103…キーボード、104…マウス、201…区間スケールバー、202…樹

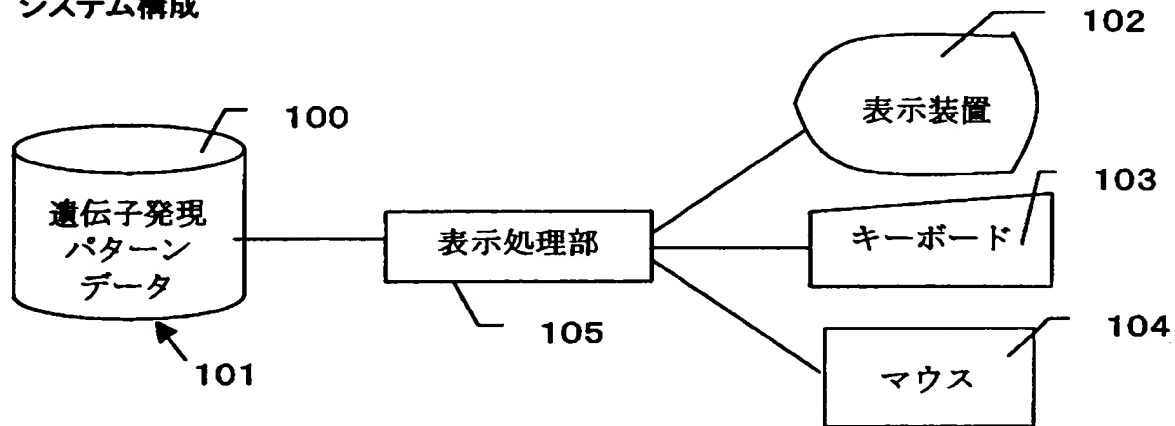
状図切断線、203…遺伝子グループ数情報、204…遺伝子グループ間境界線、205…遺伝子グループの強調表示、401…leaf型クラスタ構造体、402…node型クラス構造体。

【書類名】 図面

【図1】

# 図1

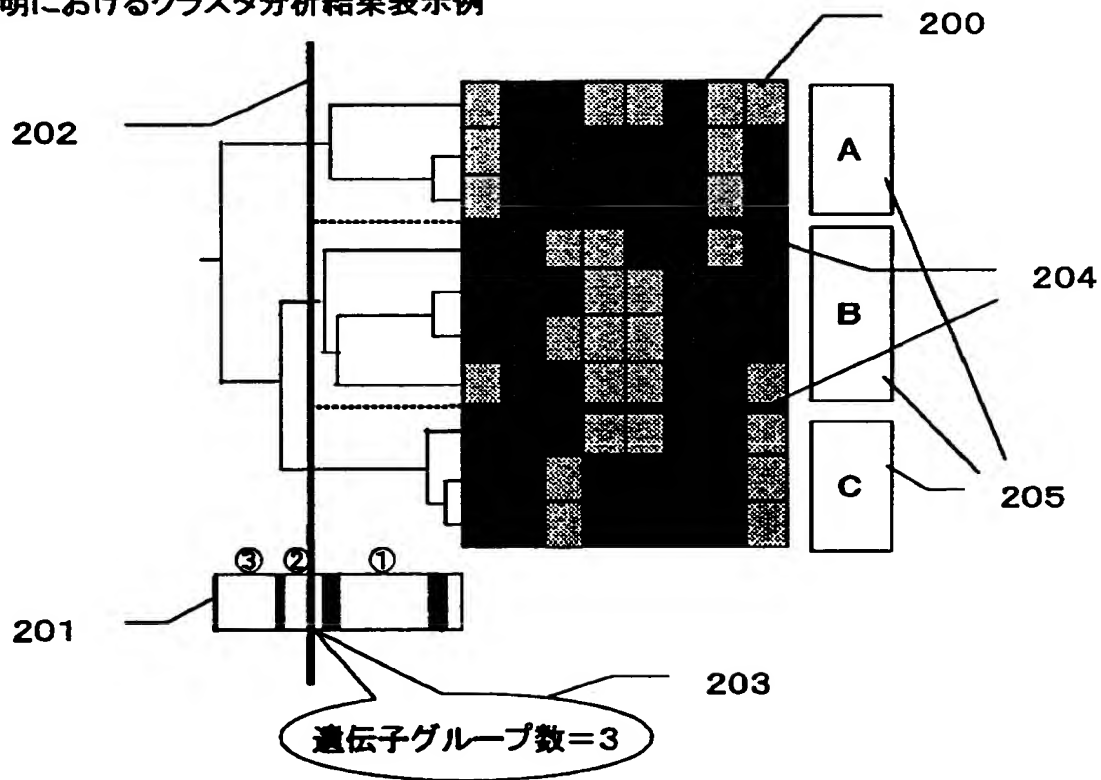
システム構成



【図2】

# 図2

本発明におけるクラスタ分析結果表示例



【図 3】

図3

遺伝子発現パターンデータ例

遺伝子ID (geneID)	実験ケース										
	1	2	3	4	5	6	..	..	..	n	
1	0	1	2	0	3	2	0				
2	1	2	0	0	2	2	1				
⋮											
⋮											
m	0	4	3	6	5	4	0				

301

302

m個のベクトルデータ

【図 4】

図4

クラスタ構造体の例

cluster		cluster	
メンバ名	値	メンバ名	値
type	leaf	type	node
left		left	● → cluster L
right		right	● → cluster R
distance		distance	91
geneID	17	geneID	
level	0	level	5

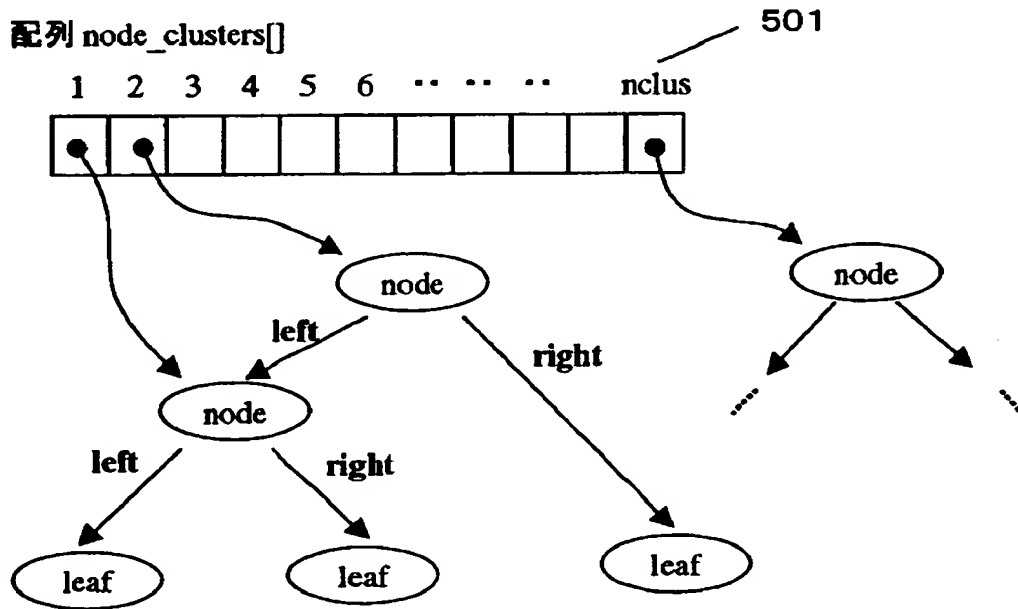
401

402

【図 5】

# 図 5

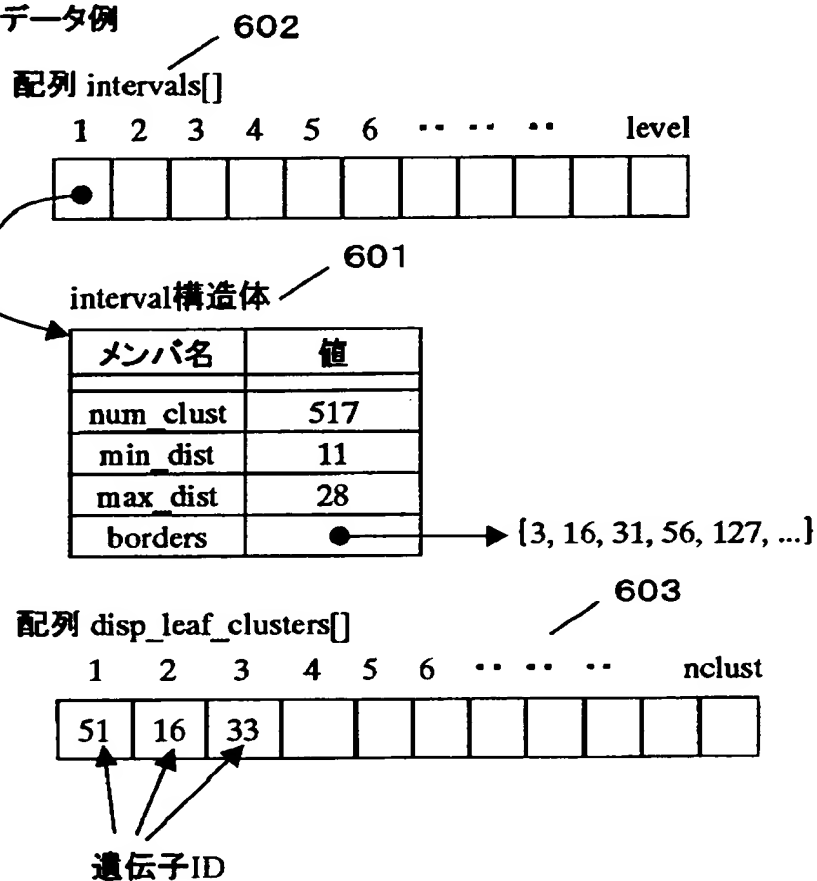
クラスタ木構造の生成例



【図 6】

図 6

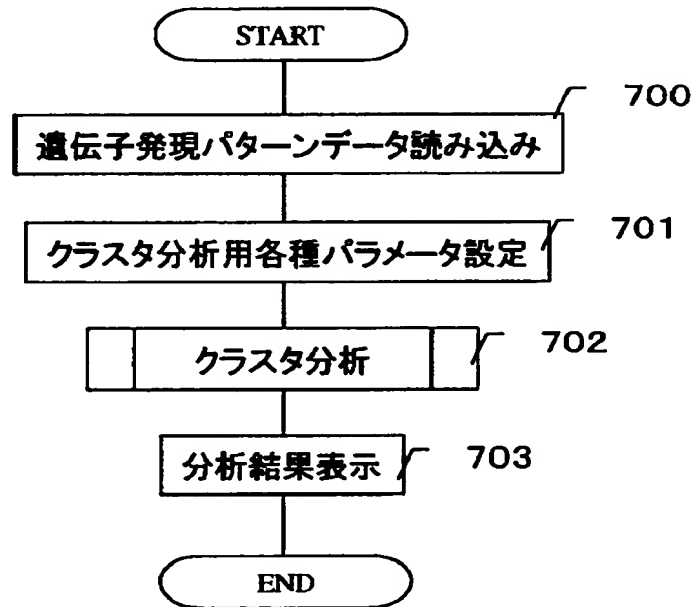
表示用データ例



【図 7】

図 7

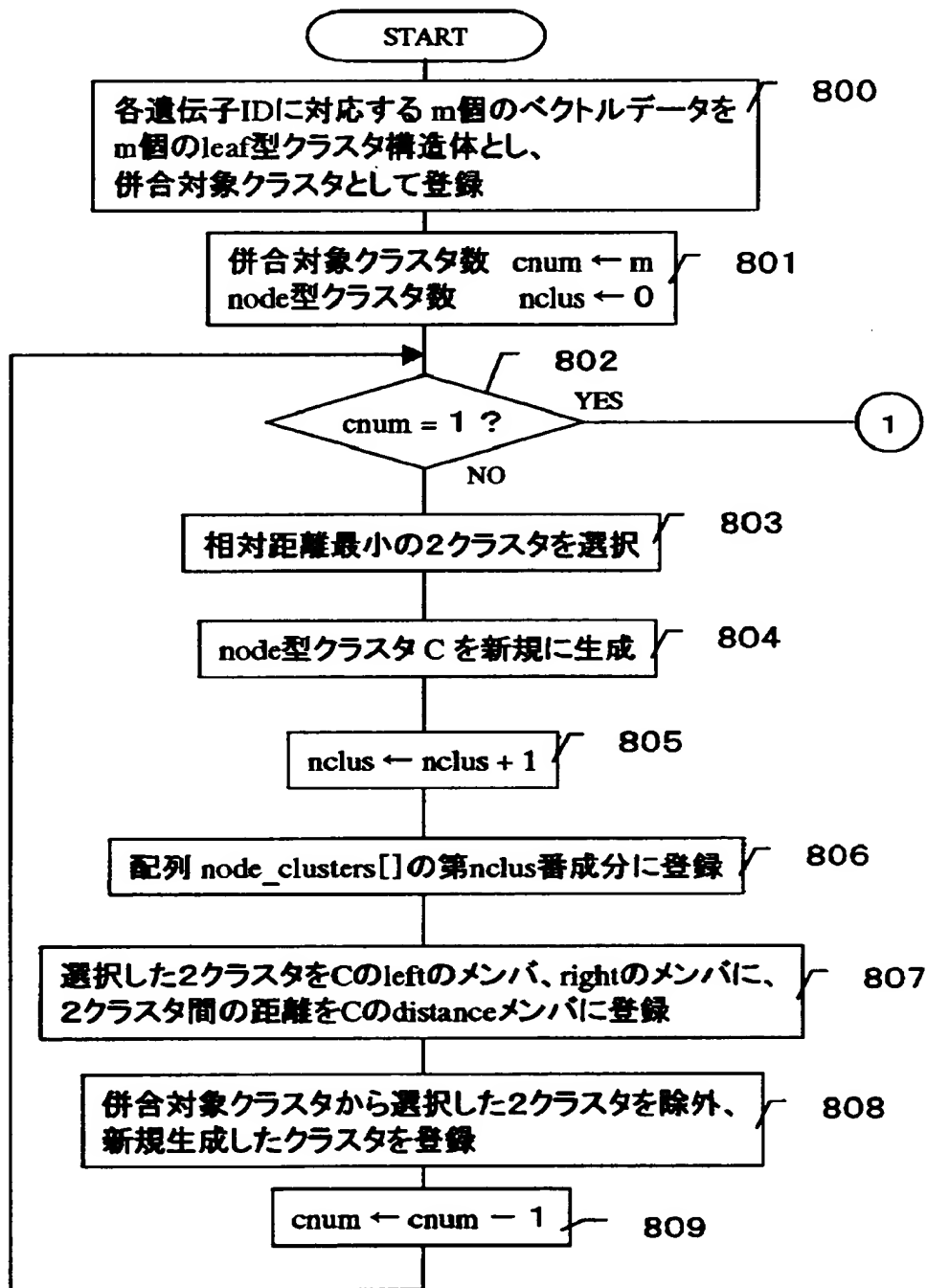
概略処理フロー



【図8】

## 図8

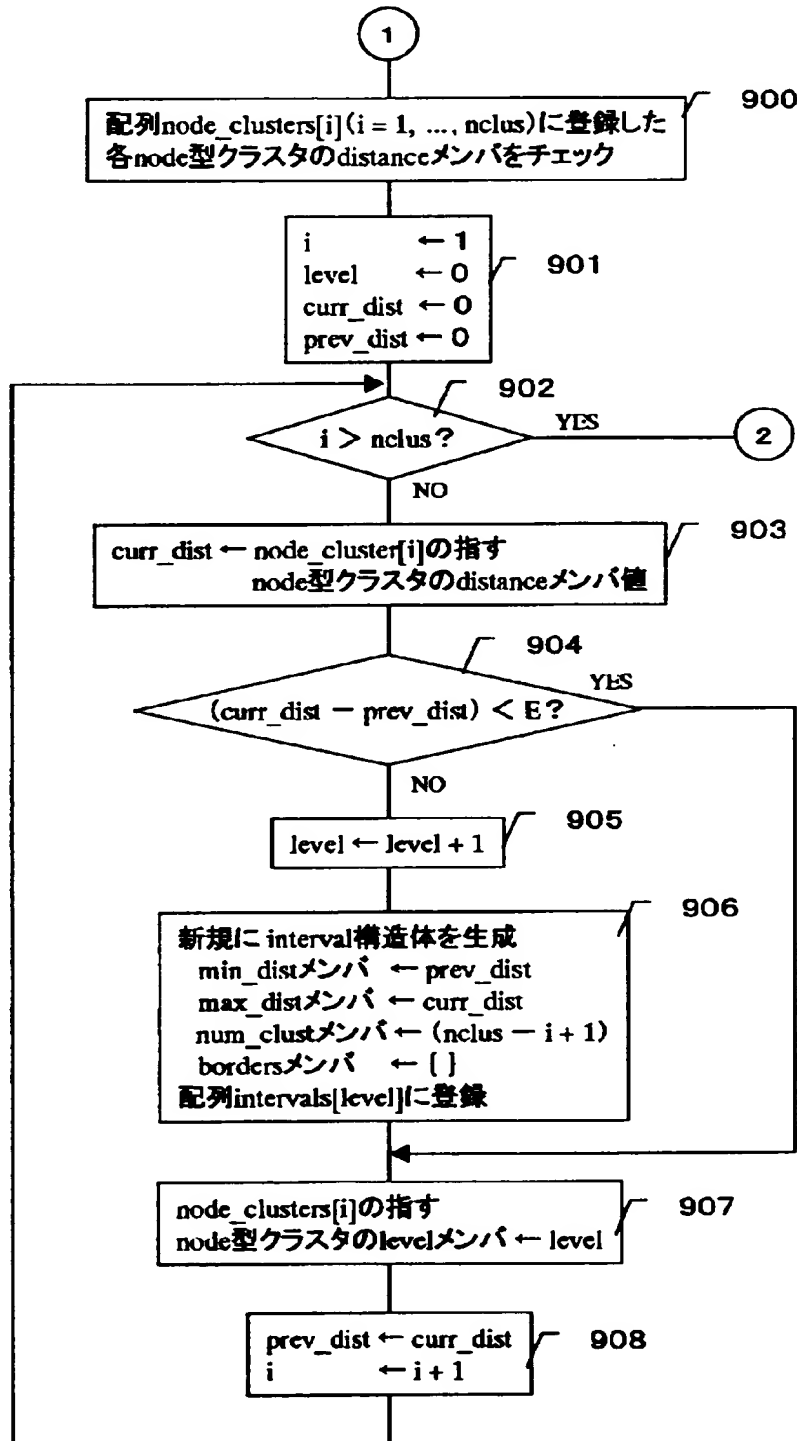
クラスタ分析の詳細フロー（その1：クラスタ木の生成）



【図 9】

図 9

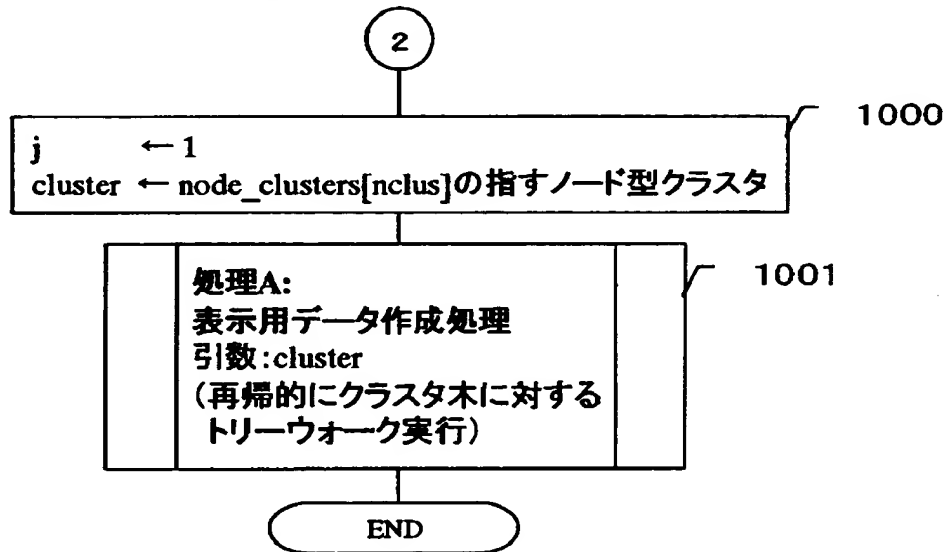
クラスタ分析の詳細フロー（その2：クラスタレベルの設定）



【図 1 0】

# 図 10

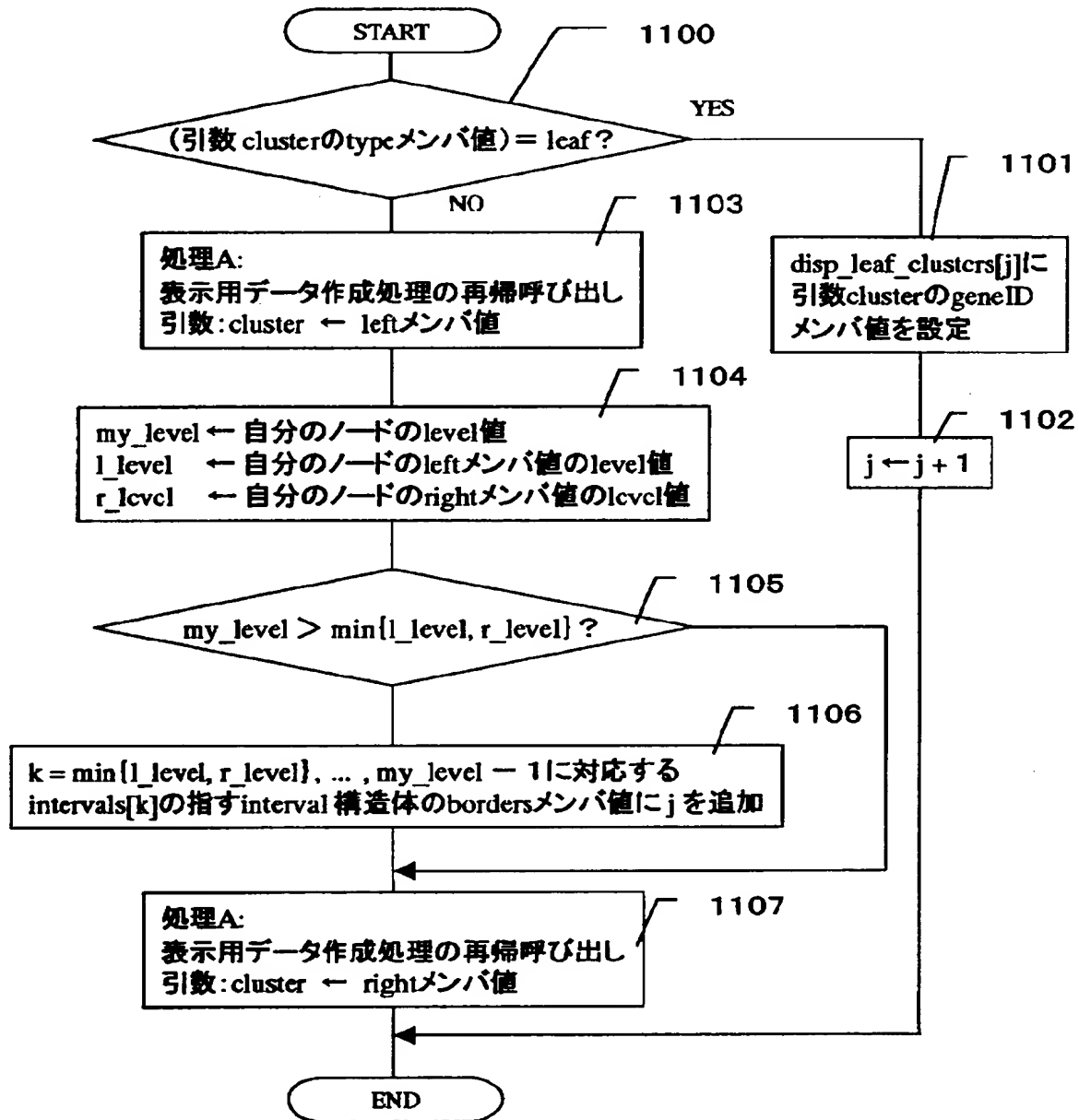
クラスタ分析の詳細フロー（その3：表示用データ作成）



【図11】

図11

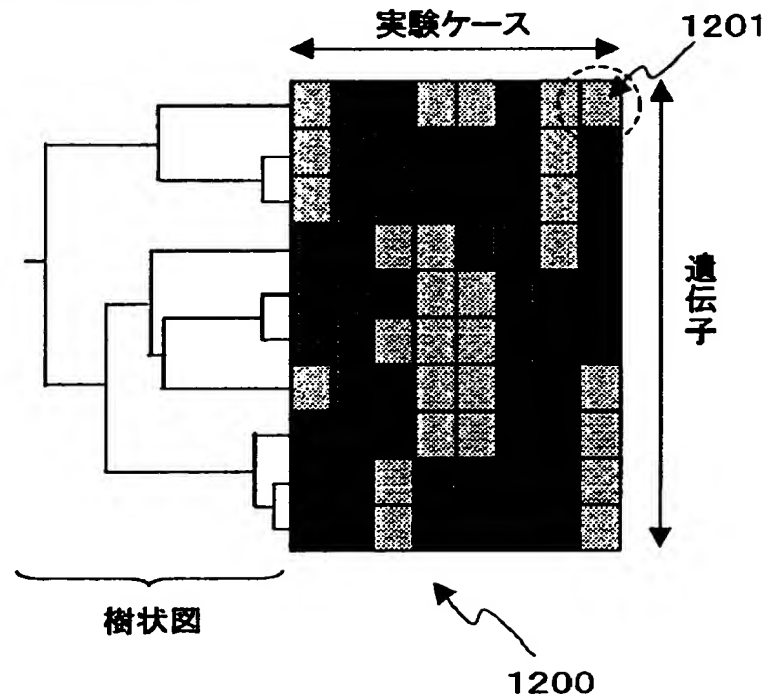
表示用データ作成処理フロー(処理A)



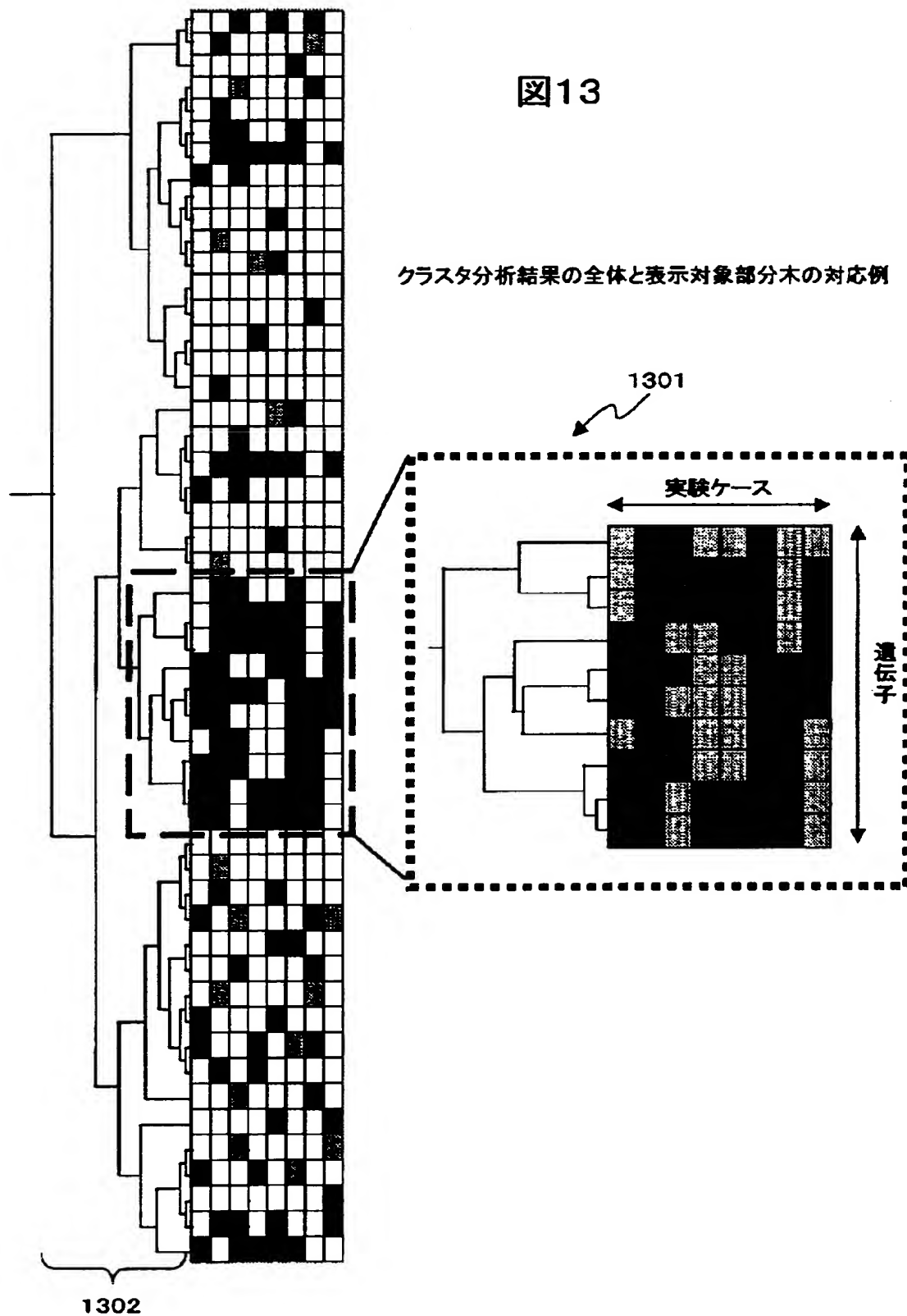
【図 1 2】

# 図 1 2

標準的クラスタ分析結果表示例



【図 1 3】



【書類名】            要約書

【要約】

【課題】    遺伝子発現パターンデータを対象としたクラスタリングの結果から、より大まかな分類結果を自動的に抽出し、ユーザが選択できるようにする。

【解決手段】    遺伝子の発現パターンのデータをクラスタ分析した結果に対し、識別誤差範囲を考慮した段階別クラスタ概数を表示する。

クラスタ分析の結果において、あるクラスタに属する遺伝子発現パターンデータの個数が所定数以上の場合に限り、遺伝子グループとしてグループ化して表示する。

【選択図】            図 2

出 願 人 履 歴 情 報

識別番号 [000233055]

1. 変更年月日	1990年 8月 7日
[変更理由]	新規登録
住 所	神奈川県横浜市中区尾上町6丁目81番地
氏 名	日立ソフトウェアエンジニアリング株式会社